【資料3】

AI事業者ガイドラインの
令和６年度更新内容

総務省
経済産業省
(令和7年3月7日･17日)

Agenda

0. AI事業者ガイドライン更新の背景

1. AI事業者ガイドラインの令和６年度の更新論点と更新方針

2. 令和６年度の更新内容

- ✓ 2-1. AIによるリスクの洗い出し・分類
- ✓ 2-2. AIの契約に関する留意事項
- ✓ 2-3. 生成AIに関する記載の追加
- ✓ 2-4. AIガバナンスに関する事例の充実
- ✓ 2-5. AIガバナンスの動向等の反映
- ✓ 2-6. 特定単語の整理・見直し
- ✓ 2-7. その他

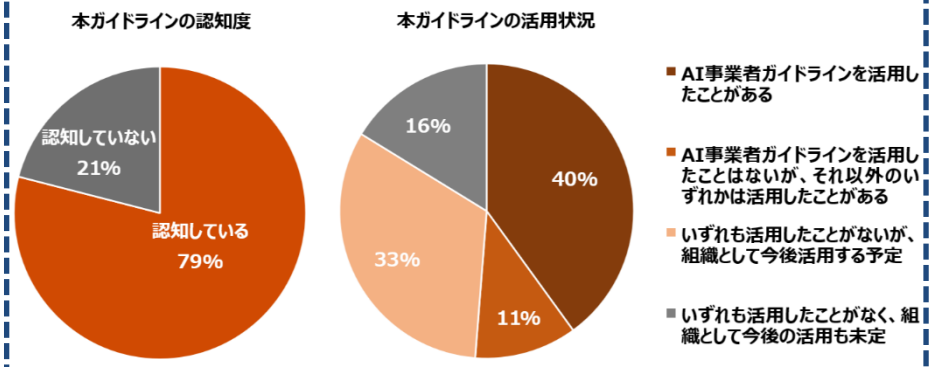
0. AI事業者ガイドライン更新の背景

- 令和6年4月に「AI事業者ガイドライン（第1.0版）」を策定・公表した後、同年11月に時点更新
- AIネットワーク社会推進会議・AIガバナンス検討会（総務省）、AI事業者ガイドライン検討会（経産省）の構成員の皆様よりいただいた御意見や、昨年末に行った両検討会における御意見のほか、事業者へのアンケート・ヒアリング、その他動向調査等を通じて、AI事業者ガイドラインの更新点となりうるポイントを抽出

構成員/委員の皆様からの主な御意見

- 生成AIの技術発展や社会浸透により、AIリスクの顕在化事例の増加や新たに考慮すべきリスク(社会変容等)が現れる中、ガイドライン上において、**AIリスクのアップデートをしてはどうか**
- **マルチモーダルAIの観点が不十分**
- RAG導入が増えているため、**RAG導入についてガイドラインでも言及すべき**
- 最新の動向である**AIエージェント**について記載すべき
- 大企業であっても生成AIの利用は始まったばかりであり、想定されるリスクの理解やガバナンス構築に向けて必要な対応の理解が難しい。**リスクベースアプローチの具体的対応事例の記載を増やしてはどうか**
- 中小企業・スタートアップの事例が少ないため、**企業規模に応じた事例**があるとよい
- 生成AIについての**責任分界の明確化**が必要

事業者へのアンケート・ヒアリング



その他動向調査等

- 国内の有識者会議や、国際動向等について調査を実施
  - ✓AI戦略会議・AI制度研究会
  - ✓文化庁「AIと著作権に関する考え方について」
  - ✓広島AIプロセス　等

- これらを踏まえ、令和６年度は次ページ記載の論点について更新（その他の論点は次年度以降に検討予定）

1. AI事業者ガイドラインの令和6年度の更新論点と更新方針

- 構成員・委員・事業者等からのご意見を踏まえ、令和６年度の更新の論点と更新方針を以下に整理

令和6年度更新論点及び更新方針 一覧

総務省検討会：AIネットワーク社会推進会議・AIガバナンス検討会
経産省検討会：AI事業者ガイドライン検討会

#	更新論点	主なご意見	更新方針
1	**AIによるリスクの洗い出し・分類**	総務省検討会	**リスクの追加・分類・マッピング** ✓ AIによる新たなリスク 及び リスクを網羅的に参照する上でのリスク分類を追加 ✓ リスク分類案とガイドラインの「共通の指針」とのマッピングを追加
2	**AIの契約に関する留意事項**	経産省検討会	**開発、提供、利用における契約に関して留意すべき事項の記載** ✓ 「契約モデルや契約当事者の多様化」に関する記載を充実化 ✓ 「開発者・提供者・利用者の間における責任分界」に関する記載を充実化
3	**生成AIに関する記載の追加**	両検討会	**生成AIに関する新たなリスクや留意点の記載** ✓ マルチモーダルな生成AIに関する記載を追加 ✓ RAG導入に関する記載を追加 ✓ プログラムコード生成に関する記載を追加 ✓ AIエージェントに関する記載を追加
4	**AIガバナンスに関する事例の充実**	両検討会 事業者	**AIガバナンスの取組事例の充実化** ✓ 「リスクベース・アプローチ」を実施する上で考慮すべき点を追加 ✓ 「グローバルなAIガバナンスを構築している企業」「中小・スタートアップ企業」「地方自治体」の事例を追加 ✓ AIガバナンスを構築する上での事業者の「人材不足」の課題を追加
5	**AIガバナンスの動向等の反映**	両検討会	**AIガバナンスに関する国内外の最新動向を追記** ✓ AI制度研究会等、国内動向において注視するべき最新状況を追記 ✓ 広島AIプロセスの動向等、国際的な動向において注視するべき最新状況を追記
6	**特定単語の整理・見直し**	両検討会	**AIガバナンスにおいて重要な単語の定義や表現の見直し** ✓ 「バイアス」の定義や表現の見直し ✓ 「透明性」の定義や表現の見直し ✓ 「多様性」「包摂性」に関する表現の揺れを修正
7	**その他**	両検討会	**UIの改善** ✓ 目次から該当ページへのリンク

1. AI事業者ガイドラインの令和6年度の更新箇所の概要

- 構成員・委員・事業者のご意見や調査結果等を踏まえ、更新箇所は以下のとおり

令和６年度の更新の論点と更新箇所 対応表

全体に関わる更新
- 3 生成AIに関する記載の追加
- 5 AIガバナンスの動向等の反映
- 6 特定単語の整理・見直し

	本編（why, what）		別添（付属資料）（how）		
主体共通	第1部 AIとは		1. 第1部関連 [AIについて]	A. AIに関する前提 B. AIによる便益/リスク	1 リスクの追加・分類・マッピング
主体共通	第2部 AIにより目指すべき社会と各主体が取り組む事項	A.「基本理念」 B.「原則」 C.「共通の指針」 D.「高度なAIシステムに関する事業者に共通の指針」 E.「AIガバナンスの構築」	2. 第2部関連 [E.AIガバナンスの構築]	A. 経営層によるAIガバナンスの構築とモニタリング B. AIガバナンスの事業者取組事例	4 AIガバナンスに関する事例の充実 4 AIガバナンスに関する事例の充実
主体別	第3部 AI開発者に関する事項	※「高度なAIシステムを開発する組織向けの広島プロセス国際行動規範」における追加的な記載事項　も含む	3. 第3部関連 [AI開発者向け]	A. 「第3部 AI開発者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説 C. 高度なAIシステムの開発にあたって遵守すべき事項	
主体別	第4部 AI提供者に関する事項		4. 第4部関連 [AI提供者向け]	A. 「第4部 AI提供者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説	
主体別	第5部 AI利用者に関する事項		5. 第5部関連 [AI利用者向け]	A. 「第5部 AI利用者に関する事項」の解説 B. 「第2部」の「共通の指針」の解説	
その他参考資料			6.「AI・データの利用に関する契約ガイドライン」を参照する際の主な留意事項について		2 AIの契約に関する留意事項
			7. チェックリスト		
			8. 主体横断的な仮想事例		
			9. 海外ガイドラインとの比較表		

4 AIガバナンスに関する事例の充実（AIガバナンス人材に関する記載の追記）

2. 令和６年度の更新内容

- 事業者がAIリスクを把握・対応しやすくなるように、「AIによるリスク」の記載を整理・拡充する

【更新内容】

① リスク事例

✓今後顕在化する可能性のある過度な依存、労働者の失業、データや利益の集中など、現状のガイドラインでカバーされていないリスク例の記載を追加

② リスクの分類

✓AIのリスクを技術的リスクと社会的リスクに大別。さらに、技術的リスクは「学習及び入力段階」「出力段階」「事後対応段階」に、社会的リスクは「倫理・法」「経済」「情報空間」「環境」の4つに分類

③ リスクと対策

✓分類した各リスクを事業者における具体対策の検討に結び付けるべく、リスクに対応する共通の指針と、事業者における対策の例を記載

【更新箇所】

✓別添1．B.AIによる便益／リスク

【更新内容の詳細】

①リスク事例

※主たる更新箇所を抜粋して表示

B.AI による便益/リスク

AI は、新規ビジネスを生み出したり、既存ビジネスの付加価値を高めたり、生産性を向上させたりする等の便益をもたらす一方で、リスクも存在する。

このリスクについては可能な限り抑制することが期待される。一方で、過度なリスク対策を講じることは、コスト増になる等、AI 活用によって得られる便益を阻害してしまうことから、リスク対策の程度をリスクの性質及び蓋然性の高さに対応させるリスクベースアプローチの考え方が重要である。

⋮

過度な依存

- 人材採用活動等、重要な意思決定を行う場面において AI による判断をそのまま用いるなど AI の不適切かつ過剰な使用により、企業が説明責任を問われたり批判を受けたりする事例が発生している。また、生成 AI を用いたチャットボットサービスと会話をしていた利用者が、AI に対し精神的な依存状態となった事例が報告されている

⋮

労働者の失業

- 生成 AI・AI 等の新たなテクノロジーは、仕事の内容（タスク）を変化させ、労働者の役割を変化させることが想定される。生成 AI・AI 等の導入により、労働者の業務負担の軽減や、労働生産性の向上が期待できる一方、失業リスクや格差の拡大なども一部では懸念されている[12]

データや利益の集中

- 一部の AI 開発者のみにデータや利益が集中する、少数言語国では自国言語による高性能な AI が存在しないといった課題も指摘されている[13]

2. 令和６年度の更新内容

【更新内容の詳細】
②リスクの分類

表 3. AI によるリスク例の体系的な分類案

・下表はAIのリスクを網羅したものではなく、想定に基づく事案も含んでおり、あくまで一例として認識することが期待される
・下表には政府等の公的機関も含めた社会全体での対応・議論が必要となるリスクも含まれる

大分類	中分類	リスク例
技術的リスク（＝主にAIシステム特有のもの）	学習及び入力段階のリスク	データ汚染攻撃等のAIシステムへの攻撃
	出力段階のリスク	バイアスのある出力、差別的出力、一貫性のない出力等
		ハルシネーション等による誤った出力
	事後対応段階のリスク	ブラックボックス化、判断に関する説明の不足
社会的リスク（＝既存のリスクがAIにおいても発生又はAIによって増幅するもの）	倫理・法に関するリスク	個人情報の不適切な取扱い
		生命等に関わる事故の発生
		トリアージにおける差別
		過度な依存
		悪用
	経済活動に関するリスク	知的財産権等の侵害
		金銭的損失
		機密情報の流出
		労働者の失業
		データや利益の集中
		資格等の侵害
	情報空間に関するリスク	偽・誤情報等の流通・拡散
		民主主義への悪影響
		フィルターバブル及びエコーチェンバー現象
		多様性・包摂性の喪失
		バイアス等の再生成
	環境に関するリスク	エネルギー使用量及び環境の負荷

2. 令和６年度の更新内容

【更新内容の詳細】
③リスクと対策

※「技術的リスク」を抜粋して表示

表 4. AI によるリスク例と共通の指針及び主体毎に重要となる事項のマッピング

・下表はAIのリスクを網羅したものではなく、想定に基づく事案も含んでおり、あくまで一例として認識することが期待される
・下表には政府等の公的機関も含めた社会全体での対応・議論が必要となるリスクも含まれる

	リスク例	関連する共通の指針	「共通の指針」に加えて主体毎に重要となる事項 第3部 AI開発者	第4部 AI提供者	第5部 AI利用者
技術的リスク	データ汚染攻撃等のAIシステムへの攻撃	5）セキュリティ確保	i. セキュリティ対策のための仕組みの導入 ii. 最新動向への留意	i. セキュリティ対策のための仕組みの導入 ii. 脆弱性への対応	i. セキュリティ対策の実施
	バイアスのある出力、差別的出力、一貫性のない出力等	1）人間中心 ①人間の尊厳及び個人の自律 ③偽情報等への対策			
	ハルシネーション等による誤った出力	2）安全性	i. 適切なデータの学習 ii. 人間の生命・身体・財産、精神及び環境に配慮した開発 iii. 適正利用に資する開発	i. 人間の生命・身体・財産、精神及び環境に配慮したリスク対策 ii. 適正利用に資する提供	i. 安全を考慮した適正利用
		3）公平性	i. データに含まれるバイアスへの配慮 ii. AIモデルのアルゴリズム等に含まれるバイアスへの配慮	i. AIシステム・サービスの構成及びデータに含まれるバイアスへの配慮	i. 入力データ又はプロンプトに含まれるバイアスへの配慮
		8）教育・リテラシー			
	ブラックボックス化、判断に関する説明の不足	6）透明性	i. 検証可能性の確保 ii. 関連するステークホルダーへの情報提供	i. システムアーキテクチャ等の文書化 ii. 関連するステークホルダーへの情報提供	i. 関連するステークホルダーへの情報提供
		7）アカウンタビリティ	i. AI提供者への「共通の指針」の	i. AI利用者への「共通の指針」の	i. 関連するステークホルダーへの

2. 令和６年度の更新内容

- AI技術を用いたサービスが普及し、契約類型が多様化する中で、典型的な契約類型を整理するとともに、経済産業省にて令和7年2月に公開した「AIの利用・開発に関する契約チェックリスト」を紹介する

【更新内容】

① 契約類型が多様化した背景

✓ AI技術を用いたサービスが普及する中で契約類型が多様化した状況について、具体的な契約場面例を追記する

② 契約類型

✓ AIの利用開発に関する典型的な３類型（汎用的AIサービス利用型、カスタマイズ型、新規開発型）を追記する

③ 市場環境の変化に応じた留意点

✓ 経済産業省「AIの利用・開発に関する契約チェックリスト」（2025年2月公表）を参照する旨追記

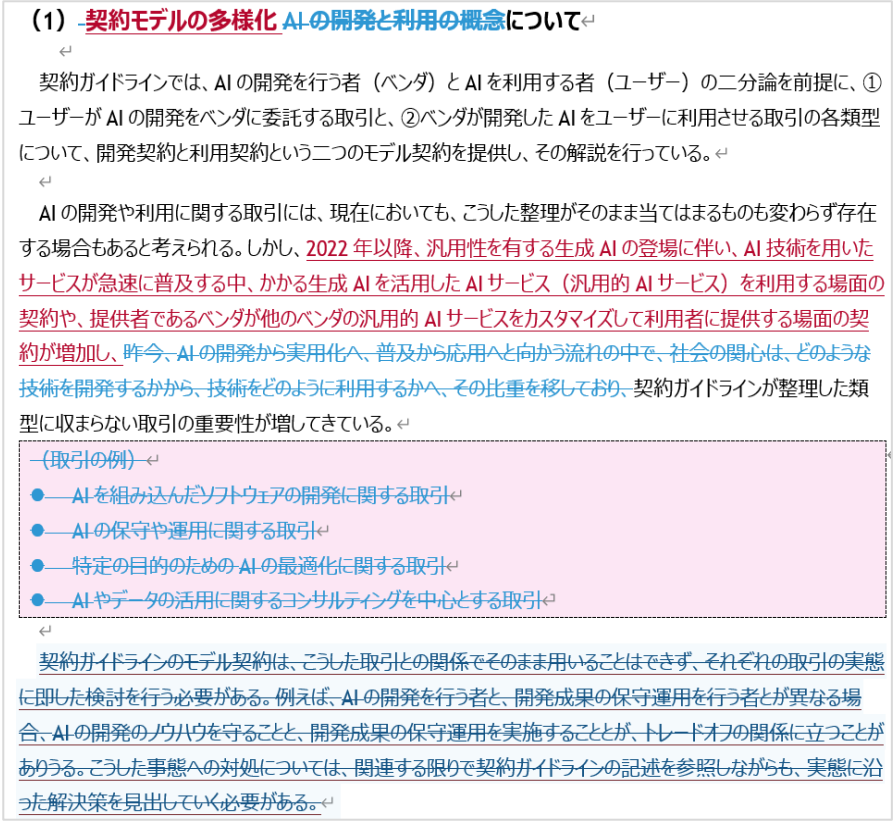
✓ 今後、契約当事者が増加し更に契約関係が重層的になる可能性や、新たな契約類型が生じる可能性があることを示す

【更新箇所】

✓ 別添6.「AI・データの利用に関する契約ガイドライン」を参照する際の主な留意事項について
(1)AIの開発と利用の概念について

【更新内容の詳細】

①契約類型が多様化した背景



（1）~~契約モデルの多様化~~ ~~AI の開発と利用の概念~~について

契約ガイドラインでは、AI の開発を行う者（ベンダ）と AI を利用する者（ユーザー）の二分論を前提に、①ユーザーが AI の開発をベンダに委託する取引と、②ベンダが開発した AI をユーザーに利用させる取引の各類型について、開発契約と利用契約という二つのモデル契約を提供し、その解説を行っている。

AI の開発や利用に関する取引には、現在においても、こうした整理がそのまま当てはまるものも変わらず存在する場合もあると考えられる。しかし、2022 年以降、汎用性を有する生成 AI の登場に伴い、AI 技術を用いたサービスが急速に普及する中、かかる生成 AI を活用した AI サービス（汎用的 AI サービス）を利用する場面の契約や、提供者であるベンダが他のベンダの汎用的 AI サービスをカスタマイズして利用者に提供する場面の契約が増加し、~~昨今、AI の開発から実用化へ、普及から応用へと向かう流れの中で、社会の関心は、どのような技術を開発するかから、技術をどのように利用するかへ、その比重を移しており、~~契約ガイドラインが整理した類型に収まらない取引の重要性が増してきている。

~~（取引の例）~~

- ~~AI を組み込んだソフトウェアの開発に関する取引~~
- ~~AI の保守や運用に関する取引~~
- ~~特定の目的のための AI の最適化に関する取引~~
- ~~AI やデータの活用に関するコンサルティングを中心とする取引~~

~~契約ガイドラインのモデル契約は、こうした取引との関係でそのまま用いることはできず、それぞれの取引の実態に即した検討を行う必要がある。例えば、AI の開発を行う者と、開発成果の保守運用を行う者とが異なる場合、AI の開発のノウハウを守ることと、開発成果の保守運用を実施することとが、トレードオフの関係に立つことがありうる。こうした事態への対処については、関連する限りで契約ガイドラインの記述を参照しながらも、実態に沿った解決策を見出していく必要がある。~~

②③の【更新内容の詳細】は次項に記載

2. 令和６年度の更新内容

【更新内容の詳細】（前項の【更新内容】②③に対応）
②契約類型

AI の利用・開発に関する契約は大きく以下の 3 類型に分類でき、それぞれの類型に応じて、留意するべき契約事項や交渉上の論点も異なる。

- 類型 1：汎用的 AI サービス利用型

AI 利用者が、AI 提供者が提供する汎用的 AI サービスを利用するケース。

- 類型 2：カスタマイズ型

AI 利用者が、AI 提供者が特定の AI 利用者向けにカスタマイズした AI サービス（カスタマイズサービス）を利用するケース。カスタマイズサービスを提供する AI 提供者は、他のベンダが提供する汎用的 AI サービスに対して、独自に開発した付加的な機能（非 AI モデル）を組み合わせる。

- 類型 3：新規開発型

AI 利用者が、AI 開発者・AI 提供者と提携して独自の AI システムを開発・利用するケース。

【更新内容の詳細】
③市場環境の変化に応じた留意点

以上のような市場環境の変化に対応するべく、経済産業省は、2025 年 2 月、我が国の事業者が AI の利活用に関する契約を検討する場面を念頭に、契約において留意するべき事項を取りまとめたチェックリスト（以下、「契約チェックリスト」という。）を策定した。[120]契約ガイドライン公表後に登場した契約類型については、契約チェックリストを参照のこと。

なお、今後の更なる AI 技術の進展に伴い、契約当事者がさらに増加し、契約関係がさらに重層的なものになることも考えられるほか、新たな契約類型が生じる可能性もあり得る。その際、契約ガイドラインや契約チェックリストに示された考え方を参照しつつも、それぞれの取引の実態に即した検討を行うことが重要と考えられる。

2. 令和６年度の更新内容

- 責任分界の明確化が求められる背景や、事故発生リスクとの関係で契約上留意するべき事項について追記する

【更新内容】

① 責任分界の明確化が求められる背景

✓事故発生時にどの主体がどのような責任を負うかについて明確な基準はなく、事業者としてリスクシナリオを描けずに、AI導入を断念・躊躇する場合もあることを示す

② 新たな類型のリスクについての対応策

✓関係当事者間で協調的にリスクを分配することが重要であることや、損害保険等の保険の活用が有用なケースもあり得ることを示す

③ 事故発生リスクとの関係で契約上の留意が有益な事項

✓契約当時には想定していなかった事故が生じる可能性や、周辺環境の変化を踏まえて契約内容の見直しを行うことの重要性を示す

✓透明性を確保することが求められる一方、セキュリティや競争力の低下リスクへの配慮も必要であることを示す

【更新箇所】

✓別添6.「AI・データの利用に関する契約ガイドライン」を参照する際の主な留意事項について
(3)AIの開発・提供・利用とアカウンタビリティについて

【更新内容の詳細】

①責任分界の明確化が求められる背景
②新たな類型のリスクについての対応策

（3） ~~AIの開発・提供・利用~~責任分界とアカウンタビリティについて

上記(2)のように、複数の当事者間でのリスク分配を検討するに当たっては、新たな類型のリスクについても分析が必要となる。AIの普及や応用が進むにつれ、AIの開発・提供・利用に伴うリスクが増えるとともに、~~そうしたリスクが~~顕在化するケースも今後増えていくことが予想されるが、。そうしたリスクの中には、AIを組み込んだソフトウェアや、これをカスタマイズ利用したサービスに関連して何らかの事故が起こり、それによりAIの開発・提供・利用の当事者や第三者に損害を生じさせるリスクがある。しかしながら、現状、事故発生時にどの主体がどのような責任を負うかについて明確な基準はなく、事業者としてリスクシナリオを描けずにAI導入を断念・躊躇する場合もあり得る。

こうした新たな類型のリスクについては、いずれかの当事者に全てのリスクを負わせるという発想ではなく、関係当事者間で協調的にリスクを分配することが重要と考えられるほか、場合によっては損害保険等の保険の活用が有用なケースもあり得る。

③の【更新内容の詳細】は次項に記載

2. 令和６年度の更新内容

【更新内容の詳細】　（前項の【更新内容】③に対応）

③事故発生リスクとの関係で契約上の留意が有益な事項

事故発生リスクとの関係で契約上留意しておくことが有益な事項としては、以下が挙げられる。

- 新たな類型のリスクの整理

  国内外では、既に AI サービスの利用に伴う知的財産権侵害、個人情報保護法違反、秘密保持契約違反等の事故が生じたケースもある。こうした先例や議論の状況等を考慮しながら、AI サービスの内容や性質に応じて考え得るリスクを洗い出し、その分配（どの主体がどのような責任を負うか）を定めておくことが重要である。一方で、契約当時には想定していなかったような事故が生じることも考えられることから、周辺環境の変化なども踏まえて契約内容の見直しを適宜行うことも重要である。

- 合理的な説明の実施

  事故発生時には、事故の原因は何か（AI 開発者、AI 提供者及び AI 利用者のそれぞれの行為に起因する可能性があるほか、AI の性質上不可避的に生じるものもありうる）及び各当事者が事故を回避するために尽くすべき注意を尽くしていたかといった点が重要な論点となり、AI の開発・提供・利用の当事者には、それぞれのプロセスにどのような関与を行ったかについて、合理的な説明を行うことが求められる可能性がある。こうした説明に対する責任は、AI の開発・提供・利用のすべての当事者の間でどのような契約が締結されていたとしても、事故について一次的な責任を負う当事者に発生する可能性があるものである。契約で定めることができるのは、契約の当事者限りでの責任の分担に限定される。契約の当事者以外の者により責任追及をされた場合に、AI のバリューチェーンに連なる者はすべて、一定の説明を求められる立場に立つ可能性がある。

- 客観的な根拠の提示

  合理的な説明を行うためには、~~説明の合理性との関係で問題となるのは、~~説明の内容に加えてその客観的な根拠~~であり~~が求められ、AI の開発・提供・利用に関する契約を締結の前後で、そうした根拠を整理しておくことが期待される。契約ガイドラインでは触れていないが、別添 2 の 3.システムデザイン（AI マネジメントシステムの構築）に関する実践例を参照の上、契約締結後の対応を検討することは有益である。

  なお、客観的な根拠を提示するに際して、モニタリングに資する情報を開示するというアプローチ（ログデータ等の開示）も考えられるが、そうした内部資料の開示は、セキュリティや競争力の低下リスクとトレードオフの関係に立ち得ることにも留意が必要である。

~~なお、AI に関する技術の進展や普及は目覚ましく、新たな技術や利用方法が日々生み出されており、それに伴い、契約において留意すべき点も変化している。契約を考える上で重要なことは、AI の開発・提供・利用のそれぞれの取引の実態に即した契約のあり方やリスクの検討を行うことであり、上述の留意事項も踏まえ、「AI・データの利用に関する契約ガイドライン」を参照されたい。~~

2. 令和６年度の更新内容

- 生成AIに関する技術の進歩や事業者における導入の進展を踏まえ、マルチモーダルな生成AI、RAG、プログラムコードの生成等に関する記載を拡充する

【更新内容】

① 生成AIによる便益を追加

✓「生成AIによる可能性」として、**マルチモーダルな生成AI**、**RAG、プログラムコードの生成、AIエージェント**によるリスクの低減やAIそのものの活用範囲の拡大といった便益に関する記載を追加

② 生成AIによるリスクを追加

✓環境への負荷、生成物による知的財産権の侵害の可能性といった生成AIの特徴を踏まえ、本編「共通の指針」の「人間中心」「安全性」等における記載を拡充

③ 生成AIに関して配慮すべき事項を追加

✓各主体向けに、マルチモーダルな生成AIやRAG、プログラムコードの生成時に配慮すべき事項（RAGにおける参照データの適切な取扱いなど）を整理し記載を追加

【主な更新箇所】

✓①別添1. B.AIによる便益／リスク
✓②本編第2部 C.共通の指針
✓③別添3.AI開発者向け
　別添4.AI提供者向け
　別添5.AI利用者向け

【更新内容の詳細】

①生成AIによる便益を追加

生成 AI による可能性

上記に加え、直近では生成 AI が台頭している。生成 AI は DX への遅れをとった日本企業の巻き返しの引き金となる可能性も高い。

日本企業の特徴として、良質の OT（Operational Technology）データの蓄積、きめ細やかなサービス及び作業等が挙げられる。これらを従来型の AI を活用することによって実現しようとした場合、組織及び業界横断的な OT データの活用、それらのサービス、作業等において AI を活用するためのデータインターフェースの統合、大量のデータの準備、多くのパターンを想定したシナリオ及びケース作り、それらを踏まえた開発等、多くの工数及び専門的な知識が必要であった。ここに生成 AI を活用すると、これらのシナリオ及びケース作り自体を自動化でき（自己教師あり学習）、幅広い企業の AI 活用を促進することが可能となる。実際に、小売企業のコールセンター及びセールスの対応に生成 AI で回答並びに資料を作成することにより、生産性を高めている事例もある。また、入力された問い合わせ及び顧客の要望に対し、社内のデータを参照することにより複数のパターンの回答・資料を作成するようなシステムが可能となっている。

さらに、テキスト・音声・画像・動画・センサ情報など、二つ以上の異なるモダリティ（データの種類）から情報を収集し、それらを統合して処理することができる生成 AI（以下、マルチモーダルな生成 AI）が台頭している。マルチモーダルな生成 AI の登場により、医療・創薬・教育・エンターテインメントなど AI の活用範囲が広がることや、推論・分析など生成以外の一般的なタスクにおける処理能力も向上すること等が期待されている。

また、RAG（検索拡張生成）の活用も広がっている。LLM による言語生成に外部情報の検索を組み合わせることによるハルシネーションの抑制、参照する情報源を指定した検索や出力文における参照元の明示等が可能となることによる出力過程・根拠の透明性向上、通常のファインチューニングと異なりモデルの再トレーニングを行うことなく参照するデータソースを追加できることによるコストの低減等が期待されている。

生成 AI のプログラムコード生成への活用も進んでいる。低コストかつ迅速なコードの生成が可能となることや、ヒューマンエラーを回避できること、高度な技能・知識がなくてもプログラミングを行えるようになること等が期待されている。

加えて、自律的な AI システム（以下、AI エージェント）も登場している。従来型の AI や生成 AI に比べより高度な効率化や自動化が可能となることで、生産性の向上につながること等が期待されている。

グローバルの激しい競争を勝ち抜くためにも、生成 AI を積極的に取り入れる形でデジタル戦略の見直しを行う等、自身が享受できる便益を正しく理解し、可能性を模索するとともに、積極的な姿勢を持つことが期待される。

【更新内容の詳細】
②生成AIによるリスクを追加

※主たる更新箇所を抜粋して表示

1) 人間中心

各主体は、AI システム・サービスの開発・提供・利用において、後述する各事項を含む全ての取り組むべき事項が導出される土台として、少なくとも憲法が保障する又は国際的に認められた人権を侵すことがないようにすべきである。また、AI が人々の能力を拡張し、多様な人々の多様な幸せ（well-being）の追求が可能となるように行動することが重要である。

① 人間の尊厳及び個人の自律

- ✧ AI が活用される際の社会的文脈を踏まえ、人間の尊厳及び個人の自律を尊重する
- ✧ 特に、AI を人間の脳・身体と連携させる場合には、その周辺技術に関する情報を踏まえつつ、諸外国及び研究機関における生命倫理の議論等を参照する

⋮

⑥ 持続可能性の確保

- ✧ AI システム・サービスの開発・提供・利用において、ライフサイクル全体で、地球環境への影響も検討する（特に、生成 AI など計算量の多い AI システムにおいては、モデルの軽量化や目的に応じたモデルの使い分け等の対策を講じる）

2) 安全性

各主体は、AI システム・サービスの開発・提供・利用を通じ、ステークホルダーの生命・身体・財産に危害を及ぼすことがないようにすべきである。加えて、精神及び環境に危害を及ぼすことがないようにすることが重要である。

① 人間の生命・身体・財産、精神及び環境への配慮

- ✧ AI システム・サービスの出力の正確性を含め、要求に対して十分に動作している（信頼性）
- ✧ 様々な状況下でパフォーマンスレベルを維持し、無関係な事象に対して著しく誤った判断を発生させないようにする（堅牢性（robustness））
- ✧ AI の活用又は意図しない AI の動作によって生じうる権利侵害の重大性、侵害発生の可能性等、当該 AI の性質・用途等に照らし、必要に応じて定期的かつ客観的なモニタリング及び対処も含めて人間がコントロールできる制御可能性を確保する

⋮

② 適正利用

- ✧ 主体のコントロールが及ぶ範囲で本来の目的を逸脱した提供・利用[20]により危害が発生することを避けるべく、AI システム・サービスの開発・提供・利用を行う
- ✧ マルチモーダルな生成 AI を中心とする AI システム・サービスの生成物については、比較的容易により精巧な生成物の生成が可能となるため、その精巧さにより誤解や偏見等を助長する可能性があることや、他者の知的財産権等を侵害する可能性があること等にも留意しつつ、その利用に際し人間の判断を介在させる等[21]の対策を講じる[22]

③ 適正学習[23]

- ✧ AI システム・サービスの特性及び用途を踏まえ、学習等に用いるデータの正確性・必要な場合には最新性（データが適切であること）等を確保する
- ✧ 学習等に用いるデータの透明性の確保、法的枠組みの遵守、AI モデルの更新等を合理的な範囲で適切に実施する
- ✧ 権利侵害複製物等の違法なコンテンツ[24]や情報等を学習データに含めないこと等に留意する[25]

2. 令和６年度の更新内容

【更新内容の詳細】
③生成AIに関して配慮すべき事項を追加

※主たる更新箇所を抜粋して表示

B.本編「第 2 部」の「共通の指針」の解説

ここでは、本編「第 5 部 AI 利用者に関する事項」では触れられていないが、本編「第 2 部」の「共通の指針」のうち、AI 利用者にとって特に重要なものについて、具体的な手法を解説する

[本編の記載内容 （再掲）] ※ 柱書のみ抜粋

1） 人間中心

各主体は、AI システム・サービスの開発・提供・利用において、後述する各事項を含む全ての取り組むべき事項が導出される土台として、少なくとも憲法が保障する又は国際的に認められた人権を侵すことがないようにすべきである。また、AI が人々の能力を拡張し、多様な人々の多様な幸せ（well-being）の追求が可能となるように行動することが重要である。

- 「⑥持続可能性の確保」関連

  [関連する記載内容]

  - AI システム・サービスの開発・提供・利用において、ライフサイクル全体で、地球環境への影響も検討する

  [具体的な手法]

  - 目的に応じて環境への影響が小さい軽量なモデルを活用するなど、モデルを適切に使い分ける

[参考文献]

- 総務省「令和 3 年版 情報通信白書」（2021 年 7 月）

2. 令和６年度の更新内容

- AI事業者ガイドラインを活用し、グローバルなAIガバナンスを構築している企業の事例として「NTT DATA」の事例を新たに追加する

【更新内容】

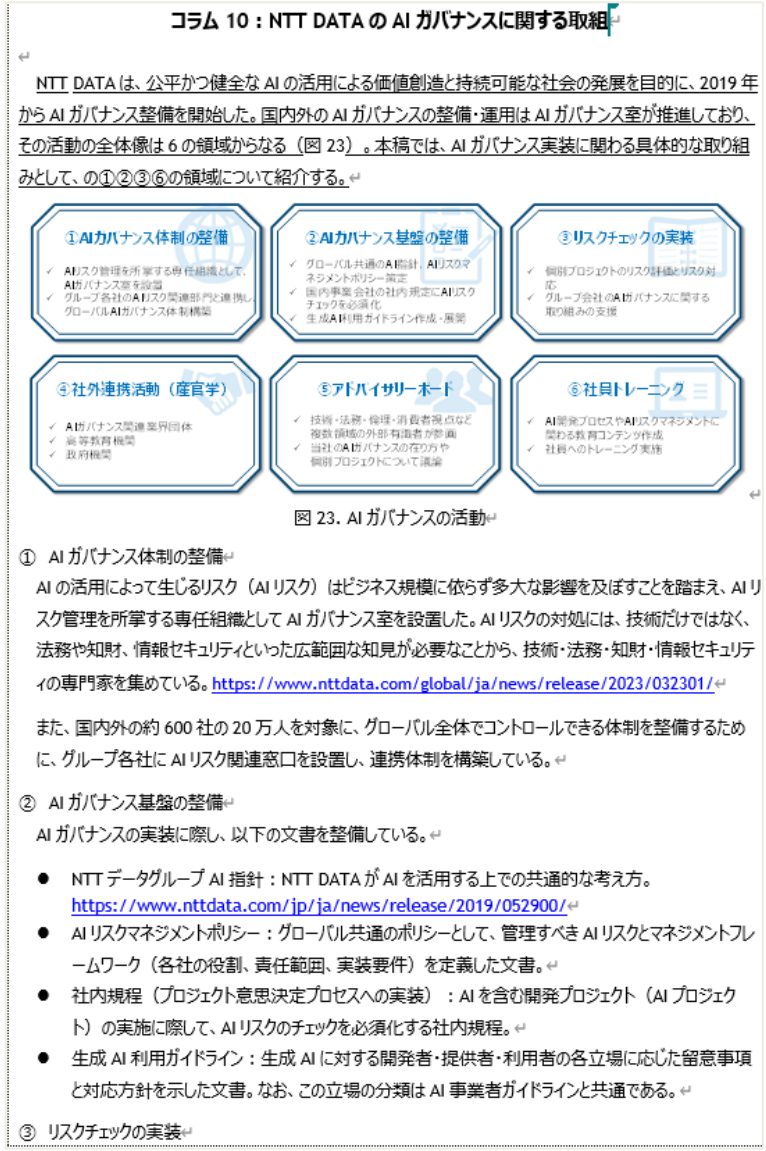
NTT DATAの取組事例をコラムとして追加

- **AIガバナンス体制の整備**
  - ✓AIリスク管理を所掌する専任組織としてAIガバナンス室を設置
  - ✓グループ各社にAIリスク関連窓口を設置し連携体制を構築
- **AIガバナンス基盤の整備**
  - ✓NTTデータグループAI指針、AIリスクマネジメントポリシー、社内規定、生成AI利用ガイドラインの整備
- **リスクチェックの実装**
  - ✓リスクの評価と対応支援の2つのステップで実装
- **社員トレーニング**
  - ✓AIリスクマネジメントやAI開発プロセスに関わる教育コンテンツを作成し社員に提供

【更新箇所】

✓別添2.「第2部E.AIガバナンスの構築」関連
B.AIガバナンスの構築に関する実際の取組事例

【更新内容の詳細】



コラム 10：NTT DATA の AI ガバナンスに関する取組

NTT DATA は、公平かつ健全な AI の活用による価値創造と持続可能な社会の発展を目的に、2019 年から AI ガバナンス整備を開始した。国内外の AI ガバナンスの整備・運用は AI ガバナンス室が推進しており、その活動の全体像は 6 の領域からなる（図 23）。本稿では、AI ガバナンス実装に関わる具体的な取り組みとして、の①②③⑥の領域について紹介する。

図 23. AI ガバナンスの活動

① AI ガバナンス体制の整備

AI の活用によって生じるリスク（AI リスク）はビジネス規模に依らず多大な影響を及ぼすことを踏まえ、AI リスク管理を所掌する専任組織として AI ガバナンス室を設置した。AI リスクの対処には、技術だけではなく、法務や知財、情報セキュリティといった広範囲な知見が必要なことから、技術・法務・知財・情報セキュリティの専門家を集めている。https://www.nttdata.com/global/ja/news/release/2023/032301/

また、国内外の約 600 社の 20 万人を対象に、グローバル全体でコントロールできる体制を整備するために、グループ各社に AI リスク関連窓口を設置し、連携体制を構築している。

② AI ガバナンス基盤の整備

AI ガバナンスの実装に際し、以下の文書を整備している。

- NTT データグループ AI 指針：NTT DATA が AI を活用する上での共通的な考え方。https://www.nttdata.com/jp/ja/news/release/2019/052900/
- AI リスクマネジメントポリシー：グローバル共通のポリシーとして、管理すべき AI リスクとマネジメントフレームワーク（各社の役割、責任範囲、実装要件）を定義した文書。
- 社内規程（プロジェクト意思決定プロセスへの実装）：AI を含む開発プロジェクト（AI プロジェクト）の実施に際して、AI リスクのチェックを必須化する社内規程。
- 生成 AI 利用ガイドライン：生成 AI に対する開発者・提供者・利用者の各立場に応じた留意事項と対応方針を示した文書。なお、この立場の分類は AI 事業者ガイドラインと共通である。

③ リスクチェックの実装

2. 令和６年度の更新内容

- 多様なステークホルダーにおける取組事例の参考として、「中小・スタートアップ企業」の事例を追加する

【更新内容】

Ubie株式会社の取組事例をコラムとして追加

① AIガバナンス体制と取組

✓ 人的資源が限定的である中で、目まぐるしい変化に対応するための利活用推進組織とリスク評価組織の連携を中心としたガバナンス体制構築

② 具体的なリスク対応の例示

✓ リスク評価における観点や主体
✓ AI提供者として留意している事柄

③ 業界団体との連携による情報取得、ルール策定への関与

✓ 業界団体を通じて、多様な企業と生成AIに関する最新情報の共有・政策動向をキャッチアップ
✓ 業界のガイドライン策定への関与

【更新箇所】

✓ 別添2.「第2部E.AIガバナンスの構築」関連
B.AIガバナンスの構築に関する実際の取組事例

【更新内容の詳細】

コラム 11： Ubie 株式会社の AI ガバナンスに関する取組

Ubie 株式会社は「テクノロジーで人々を適切な医療に案内する」ことをミッションとした医師とエンジニアで 2017 年に創業された医療 AI スタートアップ企業であり、医療機関向けの AI 問診・生成 AI サービス、生活者向けの症状検索エンジン等を提供している。

同社の AI ガバナンス体制として、社内に「生成 AI 活用推進チーム」及び「リスク・コンプライアンス委員会」を設置し、生成 AI を活用したプロダクト価値最大化や社内利用による生産性向上に取り組むと同時に、AI 利活用に当たっての法的論点やセキュリティ論点をはじめとしたリスクの検討を推進する体制を整備している（図 25.Ubie 株式会社における AI ガバナンス体制）。スタートアップ企業ゆえそれぞれの専門性を有する人材を潤沢に確保することは困難であるものの、階層型ではなくフラットな組織構造を採用することで、両会議体間における定期的な情報共有や議論を促進しており、このように縦割り意識を排して連携を行うことで、限られた人的資源の環境下でも、AI を取り巻く技術や制度に関する動向の目まぐるしい変化に応じたスピーディかつ適切な対応が可能となっている。

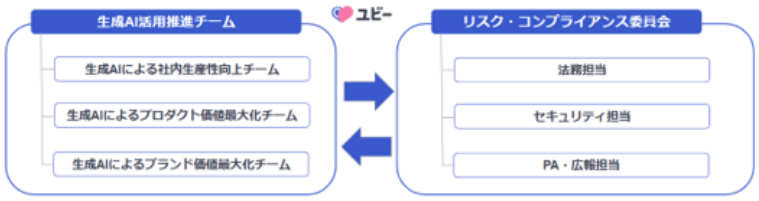


図 25.Ubie 株式会社における AI ガバナンス体制

AI リスク対応として、ベンダーリスクやクラウドセキュリティリスク、データセキュリティリスク、社内広報ガイドラインを考慮し、法務・セキュリティ・レピュテーション等の観点でリスク評価・レビューを行い、リスクの発生可能性及び影響度を評価した上で優先順位を決めつつリスク対応策を検討実施している。当該評価は客観的なものとなるよう各リスク領域の専門人材が生成 AI 開発・利用部門とは独立した立場で実施し、リスク懸念がある場合はリスク・コンプライアンス委員会にて審議する体制としている。同社では、アジャイル開発の実態に合わせたリスク評価フレームワークの型化は現在取組中である。一方、顧客向け・自社向けシステム問わず、ユースケースやデータの機密性に応じて許容できるリスクを定めリスクへの対応策を実施している。例えば、同社が生成 AI サービスを提供する際、顧客である医療機関が保有する機密性の高いデータが、AI 開発者に閲覧・利用されてしまうリスクがあるが、同社では、原則医療機関のみがデータへのアクセス権限を有する仕組みや、モデルの学習へのデータ利用を防ぐ仕組みが、オプションとして用意されている生成 AI モデルのみを AI 開発者から採用する方針としている。

さらに、これらのリスク対応方針については会社のスタンスとして社外向けに発信も行っており、同社ホームページ上にはプライバシー保護やセキュリティ確保のための取組に関する情報をまとめた「安心安全のために」ページを公開し、顧客のプライバシー保護を最も重要な経営課題の一つと位置づけ、組織全体でプライバシー課題に取り組む体制を表明している。

2. 令和６年度の更新内容

- 多様なステークホルダーにおける取組事例の参考として、「地方自治体」の事例を追加する

【更新内容】
神戸市の取組事例をコラムとして追加

- ✓ AIの利活用を推進するために条例を整備
- ✓ AI条例に基づいた、AIの活用などに関する指針を策定
- ✓ AI活用の評価を行う際、10個のリスクアセスメント項目を基に、リスクの大小（市民への影響を及ぼすかどうか）を考慮したリスク評価を実施
- ✓ 市職員の生成AIを利活用を進めるために、「神戸市生成AIの利用ガイドライン」を策定
- ✓ 条例やガイドライン等の理解浸透の実施

【更新箇所】

- ✓ 別添2.「第2部E.AIガバナンスの構築」関連
  B.AIガバナンスの構築に関する実際の取組事例

【更新内容の詳細】

コラム 12：神戸市における AI 活用のためのルール整備

神戸市は、政令指定都市であり、人口約 150 万人（政令市 7 位）、市職員数は約 2 万人（教員含む）。同市は、一定のルール下で AI を効果的かつ安全に活用することを目的として、「神戸市における AI の活用等に関する条例」（以下「AI 条例」という）を制定している。対象は、神戸市及び市の業務を請負・受託等する事業者となる。AI 条例を策定する際には、AI 事業者ガイドライン及び EU の「AI 法（Artificial Intelligence Act）」をもとに、同市が AI 利用者として果たすべき責務を盛り込んだ。ガイドラインや法規制等の単語をそのまま利用するのではなく、市職員が理解しやすい単語に置き換えて利用する等の工夫を行っている。

AI 条例は、市が行政処分等に AI を用いようとする場合にリスクアセスメントを行うことを定める。AI の持つリスクをゼロとすることは現実的ではなく、職員が AI の持つリスクを正しく認識し、そのリスクに対処する仕組みを設けることが重要である。また、「リスクベースアプローチ」の考え方を参考に、行政処分等の市民の権利利益に重大な影響を与え得る判断に AI を活用する際には、リスクに応じた慎重な手続きを課し、それ以外には簡便なチェックとすることで、AI 活用の推進と安全確保のバランスを保つことに配慮している。

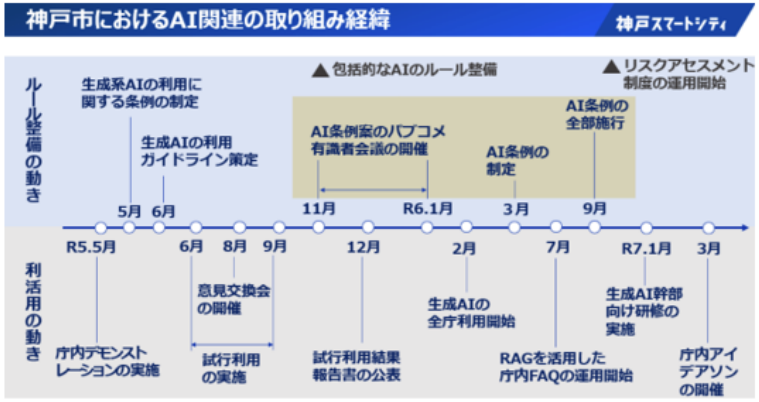


図 27. 神戸市における AI 関連の取り組み経緯

リスクアセスメントの項目として、①人間中心、②影響範囲の特定、③プライバシーの保護、④安全性の確保、⑤透明性の確保、⑥公平性の確保（バイアスへの対策）、⑦セキュリティの確保、⑧アカウンタビリティの確保、⑨職員への教育、⑩判断の責任を定めている。

リスクアセスメント手法は、リスクの大きさにより使い分けている。市民の権利・利益に重大な影響を及ぼす可能性があるものは、情報セキュリティポリシー担当部門が 48 項目のワークシートに基づき審査を行い、それ以外のものは、チェックシートに基づき所属長が確認を行う。評価基準の運用においては、専門家からなる AI 活用アドバイザーの助言も得ることとしている。

2. 令和６年度の更新内容

- 事業者から「リスクベースアプローチ」に関する課題が寄せられたことから、ヒアリング等を通じて得られた「リスクベースアプローチ」の対応を「実践例」や「実践のポイント」に追加する

【更新内容】

① リスクベースアプローチに関する実践例を追加

✓リスクベースアプローチの考え方として、透明性の確保、公平性の確保、信頼性の確保、AI利用の公表、知的財産の保護、その他の計6つの検討項目を定め、各検討項目において、潜在的なリスクと、その潜在的リスクに対するコントロール手法（最低限の対処）の評価を行う事例を追加

② リスクベースアプローチに関する実践のポイントを追加

✓ユースケース、サービスまたは製品ごとに適切なレベルの管理を実施する点を追加

【更新箇所】

✓別添2.「第2部E.AIガバナンスの構築」関連
A.経営層によるAIガバナンスの構築及びモニタリング

【更新内容の詳細】

【実践例 vi：リスクベースアプローチによるAIの提供及び利用に関する活動】

当社では、AIの提供及び利用の際のリスクベースアプローチの考え方として、透明性の確保、公平性の確保、信頼性の確保、AI利用の公表、知的財産の保護、その他の計6つの検討項目を定め、各検討項目において、潜在的なリスクと、その潜在的リスクに対するコントロール手法（最低限の対処）を定めている。例として、透明性の確保では、潜在的リスクとして、「モデルのバージョンを保存せず、AIの判断に起因する事象発生時や、検証が必要な際にAIの事後検証ができなくなるリスク」を挙げ、これに対するコントロール手法として、「AIモデル開発に用いた学習データを保管」を規定している。そして、実際のAIの提供及び利用部門が、リスクの有無とそれに対する具体的なコントロール手法の評価を行い、その結果をリスク評価部門が更に評価することで総合的にリスクへの対応の妥当性を判断している。

検討項目		潜在的なリスク	その他因有リスク①	コントロール手法	その他コントロール手法②	妥当性評価	判断
大項目	中項目		記入欄 業務担当部門		記入欄 業務担当部門	記入欄 リスク評価室	記入欄 リスク評価室
透明性の確保	データ・学習済モデルの管理	・学習・推論データ・学習済モデルのバージョンを保存せず、AIの判断に起因する事象発生時や、検証が必要な際に判断根拠の事後検証ができなくなるリスク	左記以外のリスク無し	・AIモデル開発に用いた学習データを保管 ・使用した推論データおよび推論結果の保管 … （例外規定：～であるデータは保管の対象外とする）	左記のとおり対応	①において、○○○○というリスクも考えられるのではないか	**差し戻し** ①において、考慮すべき観点の更なる検討
…	…	…	…	…	…	…	…

3.システムデザイン（AIマネジメントシステムの構築）

行動目標 3-1【ゴール及び乖離の評価、並びに乖離対応の必須化】：
各主体は、経営層のリーダーシップの下、各主体のAIのAIガバナンス・ゴールからの乖離を特定し、乖離により生じる影響を評価した上、リスクが認められる場合、その大きさ、範囲、発生頻度等を考慮して、その受容の合理性の有無を判定し、受容に合理性が認められない場合にAIの開発・提供・利用の在り方について再考を促すプロセスを、AIマネジメントシステム全体、及びAIシステム・サービスの設計段階、開発段階、利用開始前、利用開始後等の適切な段階に組み込むことが期待される。経営層は、再考プロセスについて基本方針等の方針策定、運営層はこのプロセスの具体化を行うことが重要である。そして、AIガバナンス・ゴールとの乖離評価には対象とするAIの開発・提供・利用に直接関わっていない者が加わるようにすることが期待される。なお、乖離があることのみを理由としてAIの開発・提供・利用を恣意的に不可とする対応は適当ではない。そのため、乖離評価はリスクを評価するためのステップであり、改善のためのきっかけにすぎない。

[実践のポイント]
各主体は、経営層のリーダーシップの下、以下に取り組む。
- 「AIガバナンス・ゴール」からの乖離を特定し、リスクベースアプローチを用いて、リスクに対するコントロールを選択し、ユースケース、サービス又は製品ごとに適切なレベルの管理を実施

5.評価

行動目標 5-1【AIマネジメントシステムの機能の検証】：
各主体は、経営層のリーダーシップの下、AIマネジメントシステムの設計及び運用から独立した関連する専門性を有する者に、AIガバナンス・ゴールに照らして、乖離評価プロセス等のAIマネジメントシステムが適切に設計され、適切に運用されているか否か、つまり行動目標3、4の実践を通じ、AIガバナンス・ゴールの達成に向けて、AIマネジメントシステムが適切に機能しているか否かの評価及び継続的改善を求めることが期待される。

[実践のポイント]
各主体は、経営層のリーダーシップの下、以下に取り組むことが期待される。
- 継続的改善に向けた評価の重点ポイントを、経営層が自らの言葉で明示
- AIマネジメントシステムの設計及び運用から独立した関連する専門性を有する者を割り当て
- リスクが発生する要因は変化するため、リスクベースアプローチに関して、リスクに対するコントロール、管理方法等の見直しを随時実施

2. 令和６年度の更新内容

- 事業者から「人材不足」に関する課題が寄せられたことから、ヒアリングを通じて得られた「人材不足」に関する対応を「実践のポイント」に追加する

【更新内容】

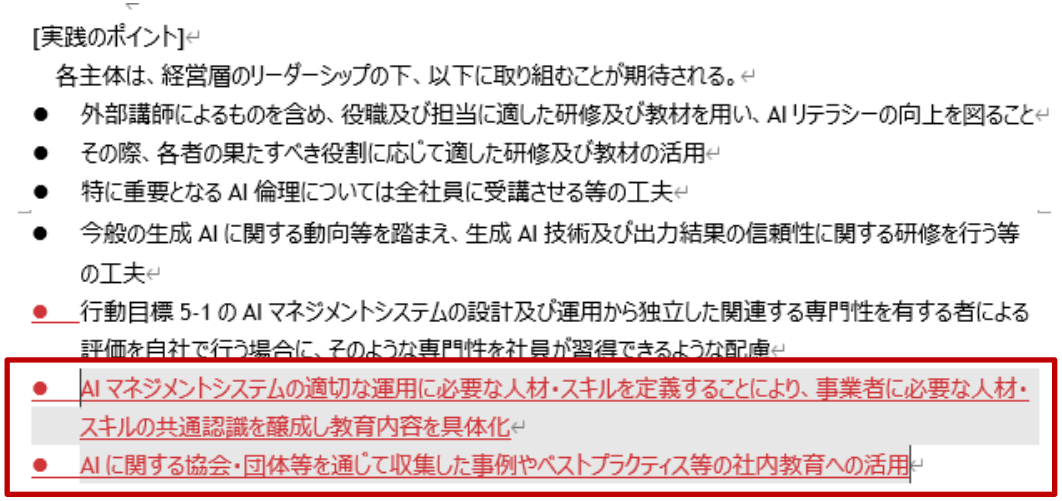
人材不足への対応策の追加

✓ AIマネジメントシステムの適切な運用に必要な人材・スキルを定義し確保
✓ AIに関する協会・団体等を通じて収集した事例やベストプラクティス等の社内教育への活用

【更新箇所】

✓ 別添2.「第2部E.AIガバナンスの構築」 関連
A.経営層によるAIガバナンスの構築及びモニタリング

【更新内容の詳細】



[実践のポイント]
各主体は、経営層のリーダーシップの下、以下に取り組むことが期待される。
- 外部講師によるものを含め、役職及び担当に適した研修及び教材を用い、AI リテラシーの向上を図ること
- その際、各者の果たすべき役割に応じて適した研修及び教材の活用
- 特に重要となる AI 倫理については全社員に受講させる等の工夫
- 今般の生成 AI に関する動向等を踏まえ、生成 AI 技術及び出力結果の信頼性に関する研修を行う等の工夫
- 行動目標 5-1 の AI マネジメントシステムの設計及び運用から独立した関連する専門性を有する者による評価を自社で行う場合に、そのような専門性を社員が習得できるような配慮
- AI マネジメントシステムの適切な運用に必要な人材・スキルを定義することにより、事業者に必要な人材・スキルの共通認識を醸成し教育内容を具体化
- AI に関する協会・団体等を通じて収集した事例やベストプラクティス等の社内教育への活用

2. 令和６年度の更新内容

- AI制度研究会等の国内動向や、広島AIプロセス等の国際動向において、注視するべき最新状況を追記する

【更新内容】

主に以下のトピックを追加

＜国内＞

① AI戦略会議・AI制度研究会 中間とりまとめ
② 内閣府「消費者をエンパワーするデジタル技術に関する専門調査会」報告書
③ 文化庁「AI と著作権に関する考え方について」
④ 個人情報保護委員会
「個人情報保護法 いわゆる3年ごと見直しに係る検討」
⑤ デジタル庁「AI時代における自動運転車の社会的ルールの在り方検討サブワーキンググループ」
⑥ 経済産業省「コンテンツ制作のための生成AI利活用ガイドブック」
⑦ 文部科学省「初等中等教育段階における生成AIの利活用に関するガイドライン」

＜国際＞

⑧ 広島AIプロセス
⑨ EUの「AI法（Artificial Intelligence Act）」

【主な更新箇所】

✓①、②：本編 はじめに
✓③、④、⑤、⑥、⑧、⑨：本編第2部 C.共通の指針
✓⑦：別添2.「第2部 E.ガバナンスの構築」関連
✓⑧：別添3.AI開発者向け

【更新内容の詳細】

①AI戦略会議・AI制度研究会 中間とりまとめ

[9] 2025 年 2 月 4 日に、AI 戦略会議・AI 制度研究会中間とりまとめが公表された。
https://www8.cao.go.jp/cstp/ai/ai_senryaku/13kai/shiryou2.pdf
2024 年 8 月 2 日から、AI 戦略会議のもとに設置された「AI 制度研究会」において、AI 関係者へのヒアリングや海外の事例研究などを通じて、AI 制度のあり方についての検討が始まっている。
https://www8.cao.go.jp/cstp/ai/ai_kenkyu/ai_kenkyu.html

※本編 はじめに 脚注9

②内閣府「消費者をエンパワーするデジタル技術に関する専門調査会」報告書

[2] 一般消費者が AI、特に生成 AI を活用するにあたっての注意点等については、消費者庁「AI 利活用ハンドブック～生成 AI 編～」（2024 年 5 月）を参照。
https://www.caa.go.jp/policies/policy/consumer_policy/information/ai_handbook
また、消費者を支援することに活用できる AI を含むデジタル技術の現状や見通し、課題等については、内閣府「消費者をエンパワーするデジタル技術に関する専門調査会」報告書で整理されている。
https://www.cao.go.jp/consumer/iinkaikouhyou/2024/doc/202412_digital_technology_houkoku.pdf

※本編 はじめに 脚注2

③文化庁「AI と著作権に関する考え方について」

③ 適正学習[23]
- ✧ AI システム・サービスの特性及び用途を踏まえ、学習等に用いるデータの正確性・必要な場合には最新性（データが適切であること）等を確保する
- ✧ 学習等に用いるデータの透明性の確保、法的枠組みの遵守、AI モデルの更新等を合理的な範囲で適切に実施する
- ✧ 権利侵害複製物等の違法なコンテンツ[24]や情報等を学習データに含めないこと等に留意する[25]

[24] 文化庁「AI と著作権に関する考え方について」（文化審議会著作権分科会法制度小委員会、2024 年 3 月）において、海賊版等、違法にアップロードされているものも学習されてしまうことが権利者の懸念として挙げられている。

※本編 共通の指針 脚注24

2. 令和６年度の更新内容

【更新内容の詳細】

④個人情報保護委員会
「個人情報保護法　いわゆる3年ごと見直しに係る検討」

[30] 個人情報保護法については、個人情報保護委員会において、いわゆる３年ごと見直しに関する検討が進められている。その中で、AI 開発等を含む統計作成等のみを目的とした取り扱いを実施する場合の本人同意の在り方等が検討されている。
https://www.ppc.go.jp/personalinfo/3nengotominaoshi/
（2025 年２月公表資料：https://www.ppc.go.jp/files/pdf/seidotekikadainitaisurukangaekatanitsuite_r6.pdf）

※本編 共通の指針 脚注30

⑤デジタル庁「AI時代における自動運転車の社会的ルールの在り方検討サブワーキンググループ」

[10] このような可能性も踏まえ、自動運転の社会実装にあたり、AI がより適切な判断を下せるようにすることなどより安全性を高める等の観点からデジタル庁において「AI 時代における自動運転車の社会的ルールの在り方検討サブワーキンググループ」が設置されている。本サブワーキンググループでは、事故発生時のデータだけではなく、それ以外の走行データ（ニアミス等）も含めて収集し、関係者で共有・分析する仕組みを構築することや、より適正なプログラム作成に資するようルールの明確化・具体化を図ること等の提言をとりまとめており、それを受けて関係省庁が検討を進めている。
https://www.digital.go.jp/councils/mobility-subworking-group

※本編 共通の指針 脚注33

⑥経済産業省「コンテンツ制作のための生成AI利活用ガイドブック」

[37] 経済産業省・IPA は、個人の学習及び企業の人材確保・育成の指針として DX 時代の人材像を「デジタルスキル標準」として整理（2022 年 12 月）。~~生成 AI の利用を通じた更なる企業 DX の推進に向けて~~、2023 年 8 月に「生成 AI 時代の DX 推進に必要な人材・スキルの考え方」を取りまとめ、指示（プロンプト）の習熟、「問いを立てる」「仮説検証する」等の必要性をスキル標準に反映し、2024 年 7 月には、普及する生成 AI の影響を踏まえ、新技術への向き合い方等を補記として追加し、「データ活用」「テクノロジー」の学習項目例に「大規模言語モデル・画像生成モデル・オーディオ生成モデル」等の生成 AI 関連の技術を追加。
・デジタルスキル標準　https://www.meti.go.jp/policy/it_policy/jinzai/skill_standard/main.html
・デジタル時代の人材政策に関する検討会　https://www.meti.go.jp/shingikai/mono_info_service/digital_jinzai/index.html
また、総務省は今後の生活の中で生成 AI に触れうる国民（初心者）向けに、生成 AI の基礎知識、生成 AI の活用場面や入門的な使い方、生成 AI 活用時の注意点などを紹介する「生成 AI はじめの一歩～生成 AI の入門的な使い方と注意点～」を公表。
https://www.soumu.go.jp/use_the_internet_wisely/special/generativeai/

※本編 共通の指針 脚注37

⑦文部科学省「初等中等教育段階における生成AIの利活用に関するガイドライン」

- 教育
  文部科学省「初等中等教育段階における生成 AI の利活用に関するガイドライン」（2024 年 12 月）[47]

※別添2.「第2部 E.ガバナンスの構築」関連

⑧広島AIプロセス

当該「行動規範」の遵守状況を高度な AI システムを開発する AI 開発者自らが自主的に確認し報告する「報告枠組み」が OECD との協力の下、G7 で合意され、2025 年 2 月より運用開始されている[46]。当該「行動規範」を遵守した高度な AI システムを開発する AI 開発者は、「報告枠組み」に参加することが期待されている。

※本編 共通の指針

「高度な AI システムを開発する組織向けの広島プロセス国際行動規範」の遵守状況を高度な AI システムを開発する AI 開発者自らが自主的に確認し報告する「報告枠組み」が OECD との協力の下、G7 で合意され、2025 年 2 月より運用開始されている[104]。当該「行動規範」を遵守した高度な AI システムを開発する AI 開発者は、「報告枠組み」に参加することが期待されている。

なお、質問項目は以下の 7 つに大別されている。

(1) リスクの特定及び評価
(2) リスク管理及び情報セキュリティ
(3) 高度な AI システムに関する透明性報告
(4) 組織の統治、インシデント管理及び透明性
(5) 内容の認証及び来歴確認の仕組み
(6) AI の安全性向上と社会リスクの軽減に向けた研究及び投資
(7) 人間と世界の利益の促進

※別添3. AI開発者向け

⑨EU AI法

[33] 透明性については、諸外国でも様々な定義がある。例えば、NIST,Artificial Intelligence Risk Management Framework（January 2023）では、透明性（システムで何が起きたかについて答えられること）、説明可能性（システムでどのように決定がなされたかについて答えられること）及び解釈可能性（なぜその決定がされたかについてその意味又は文脈について答えられること）に分類されており、European Commission, ETHICS GUIDELINES FOR TRUSTWORTHY AI （April 2019）では、トレーサビリティ、説明可能性及びコミュニケーションが取り上げられている。また、国際標準(ISO/IEC JTC1/SC42)では、透明性（適切な情報が関係者に提供されること）と定義されている。その他に EU, AI 法（Artificial Intelligence Act）（June 2024）において透明性とは、AI システムが適切な追跡可能性と説明可能性を実現する方法で開発及び利用され、人間が AI システムで通信又は対話していることを認識できるようにし、AI システムの機能と限界について利用者に適切に開示し、影響を受ける人間に対して権利について説明することを意味する。本文書では、情報開示に関する事項を広く「透明性」とする。なお定義のほか、開示対象者や開示主体、開示目的が諸外国によって異なることにも留意する。

※本編 共通の指針 脚注33

2. 令和６年度の更新内容

- 経営層のAIガバナンスの管理・監督責任が問われる可能性があることにも留意するため、経営層の取組の補足として、脚注を追加する

【更新内容】

経営層へのAIガバナンスの取組の追加

✓ 経営層のAIガバナンスの管理・監督責任が問われる場合があることにも留意

【更新箇所】

✓ 別添2.「第2部E.AIガバナンスの構築」関連
A.経営層によるAIガバナンスの構築及びモニタリング

【更新内容の詳細】

[実践のポイント]

各主体は、経営層のリーダーシップの下、以下に取り組む[18]。

- 事業における価値の創出、社会課題の解決等の AI の開発・提供・利用の目的を明確に定義
- 自社の事業に結びつく形で、「便益」及び意図せざるものを含めた「リスク」を具体的に理解
- その際に、回避すべき「リスク」及び複数主体にまたがる論点に留意し、バリューチェーン/リスクチェーン全体で便益を確保、リスクを削減
- 迅速に経営層に報告/共有する仕組みを構築

「リスク」としては、具体的には以下のようなものが挙げられ、これらのリスクに起因して、レピュテーションの低下及び法令違反を理由とした制裁金並びに損害賠償責任の負担等による損失が生じる可能性もある。リスクについての詳細は、別添 1.「B.AI による便益/リスク」を参照いただきたい。

- AI 全般に共通するリスク
  - バイアスのある結果及び差別的な結果の出力、フィルターバブル・エコーチェンバー、偽情報、不適切な個人情報の取扱い、データ汚染攻撃、ブラックボックス化、機密データの漏洩、AI システム・サービスの悪用、エネルギー使用量及び環境の負荷、バイアスの再生成　等
- 生成 AI により顕在化したリスク
  - ハルシネーション、誤情報を鵜呑みにすること、著作権等の権利及び資格との関係　等
- 組織・管理に起因するリスク
  - 製品又はサービスに AI が含まれていることの不認識、ガバナンスにおける AI に関する考慮不足、環境認識又は計画等が不足したことによる不適切、偏在的な AI の活用、仕事の棲み分け、人間と AI との間の関係性の整理不足　等

なお、バリューチェーン/リスクチェーン全体での便益の確保、リスクの削減に努めるために重要な複数主体にまたがる論点として、例えば、以下のものが挙げられる。

- 主体間、又は主体内の責任分配
- AI システム・サービス全体の品質向上
- 各 AI システム・サービスが相互に繋がることによる新たな価値の創出の可能性（System of Systems）
- AI 利用者・業務外利用者のリテラシー向上

[18] 経営層の AI ガバナンスの管理・監督責任が問われる場合があることにも留意する。

2. 令和６年度の更新内容

- 「バイアス」「多様性・包摂性」「透明性」など、AIガバナンスにおいて重要な単語の定義や表現を見直す

【更新内容】
以下の単語の定義・表現を見直し

① 「バイアス」の定義・表現の見直し
✓ ガイドラインにおける「バイアス」の定義を記載するとともに、「～なバイアス」「○○的バイアス」などの表現を見直し、明確化

② 「多様性・包摂性」の定義・表現の見直し
✓ ガイドライン内の「多様性」や「包摂性」という用語を「多様性・包摂性」に統一（表現の揺れの修正）

③ 「透明性」の定義・表現の見直し
✓ EU AI法の定義を参考として追加
✓ 諸外国の開示対象者や開示主体、開示目的が異なる等、「透明性」の確保にあたっての留意点を追加

【主な更新箇所】
✓ ①本編第2部 C.共通の指針
　　別添3.AI開発者向け
✓ ②本編第2部 C.共通の指針
✓ ③本編第2部 C.共通の指針

【更新内容の詳細】

①「バイアス」の定義・表現の見直し

[26]「バイアス」という言葉には例えば以下の通り様々な解釈が考えられ、本ガイドラインではそれらを総称したものとして用いている。
・統計的な用語（サンプリングバイアス。偏り・偏差など。）
・心理学的な用語（認知バイアス（思い込み等に起因。集団ごとの社会通念等による社会的バイアスも含む）、感情バイアス（人間の感情・都合等に起因）など。）
また NIST の Proposal for Identifying and Managing Bias in Artificial Intelligence (SP 1270) では、AI において、Systemic（既存のルールや規範、慣行等によるもの）、Statistical and Computational (統計的・計数的なもの)、Human（認知・知覚、習性等によるもの）という３つのカテゴリと、それぞれに属する典型的なバイアスがある旨が説明されている

- D-3）i. データに含まれるバイアスへの配慮
  - 学習データ、AI モデルの学習過程によってバイアス（学習データには現れない潜在的なバイアスを含む）が含まれうることに留意し、データの質を管理するための相当の措置を講じる（「3）公平性」）
  - 学習データ、AI モデルの学習過程からバイアスを完全に排除できないことを踏まえ、AI モデルが代表的なデータセットで学習され、AI システムに不公正なバイアス（特定の個人・集団に対し、合理的に説明できない不利益をもたらすバイアス）がないか点検されることを確保する（「3）公平性」）

※主たる更新箇所を抜粋して表示

2. 令和６年度の更新内容

【更新内容の詳細】
②「多様性・包摂性」の定義・表現の見直し

※主たる更新箇所を抜粋して表示

C. 共通の指針

取組にあたり、各主体は、以下に述べる「1）人間中心」に照らし、法の支配、人権、民主主義、多様性・包摂性及び公平公正な社会を尊重するよう AI システム・サービスを開発・提供・利用すべきである。また、憲法、知的財産関連法令及び個人情報保護法をはじめとする関連法令、AI に係る個別分野の既存法令等を遵守すべきであり、国際的な指針等の検討状況についても留意することが重要である[13]。

なお、これらの取組は、各主体が開発・提供・利用する AI システム・サービスの特性、用途、目的及び社会的文脈を踏まえ、各主体の資源制約を考慮しながら自主的に進めることが重要である。

【更新内容の詳細】
③「透明性」の定義・表現の見直し

[33] 透明性については、諸外国でも様々な定義がある。例えば、NIST,Artificial Intelligence Risk Management Framework（January 2023）では、透明性（システムで何が起きたかについて答えられること）、説明可能性（システムでどのように決定がなされたかについて答えられること）及び解釈可能性（なぜその決定がされたかについてその意味又は文脈について答えられること）に分類されており、European Commission, ETHICS GUIDELINES FOR TRUSTWORTHY AI （April 2019）では、トレーサビリティ、説明可能性及びコミュニケーションが取り上げられている。また、国際標準(ISO/IEC JTC1/SC42)では、透明性（適切な情報が関係者に提供されること）と定義されている。その他に EU, AI 法（Artificial Intelligence Act）（June 2024）において透明性とは、AI システムが適切な追跡可能性と説明可能性を実現する方法で開発及び利用され、人間が AI システムで通信又は対話していることを認識できるようにし、AI システムの機能と限界について利用者に適切に開示し、影響を受ける人間に対して権利について説明することを意味する。本文書では、情報開示に関する事項を広く「透明性」とする。なお定義のほか、開示対象者や開示主体、開示目的が諸外国によって異なることにも留意する。

総務省
Ministry of Internal Affairs and Communications

経済産業省
Ministry of Economy, Trade and Industry